CLAIMS

We claim:

1. A method for converting a digitized melody into a sequence of notes, comprising:

segmenting said melody into a series of frames;

computing a spectral energy distribution (SED) indicator for each frame; and

estimating initial breakpoints in said melody based on said SED indicator, said notes being defined between adjacent initial breakpoints.

2. A method according to claim 1, wherein the value of said SED indicator for a given frame is relatively large if an energy distribution associated with said frame is concentrated in one or more specified frequency bands.

3. A method according to claim 2, including filtering said melody with a high pass filter prior to segmenting said melody into said frames.

4. A method according to claim 3, wherein said energy distribution is determined from a normalized energy spectrum of said frame.

5. A method according to claim 3, wherein said specified frequency band is the upper portion of a 0 to 4 kHz range.

6. A method according to claim 3, wherein the SED indicator is defined as $\frac{\sum_k f(k)g(X(k))}{\sum_k g(X(k))}$, where $X(k)$ is the energy spectrum of a frame at frequency bin k and $f(k)$ and $g(X(k))$ are non-negative and non-decreasing functions of k and $X(k)$, respectively.

7. A method according to claim 6, wherein the SED indicator is defined as $\frac{\sum_k kX(k)}{\sum_k X(k)}$.

8. A method according to claim 6, wherein the SED indicator is defined as $\frac{\sum_k \sqrt{k}X(k)}{\sum_k X(k)}$.

9. A method according to claim 6, wherein the SED indicator is defined as $\frac{\sum_k k^2X(k)}{\sum_k X(k)}$.

10. A method according to claim 6, wherein the SED indicator is defined as $\frac{\sum_k \sin\left(\frac{\pi k}{2K}\right)X(k)}{\sum_k X(k)}$, where K is the frequency bin corresponding to the Nyquist frequency.

11. A method according to claim 6, wherein the SED indicator is defined as $\frac{\sum_k kX(k)^2}{\sum_k X(k)^2}$.

12. A method according to claim 3, wherein the auto-correlation of each said frame is computed and said SED indicator is computed by estimating the slope at the origin of the frame's auto-correlation and normalizing that slope by the value at the origin.

13. A method according to claim 1, including estimating the pitch of each said frame.

14. A method according to claim 13, wherein estimating the pitch of each frame comprises:

computing the auto-correlation of each said frame; and

estimating the pitch of each said frame by selecting a pitch period corresponding to a shift where the auto-correlation coefficient associated with the frame is relatively large.

15. A method according to claim 1, including estimating the pitch of each said note between adjacent initial breakpoints.

16. A method according to claim 15, wherein estimating the pitch of each note between initial breakpoints comprises:

computing the auto-correlation of each said frame;

estimating the pitch of each said frame by selecting a pitch period corresponding to a shift where the auto-correlation coefficient associated with the frame is relatively large; and

averaging or taking the median of the pitch estimates of frames between adjacent breakpoints.

17. A method according to claim 15, including associating each said initial breakpoint with a confidence level, which is influenced by at least one of (a) the degree in the change or rate of change of pitch in the frames around the initial breakpoints, and (b) the value of said SED indicator in the vicinity of the initial breakpoint.

18. A method according to claim 17, wherein the confidence level is further influenced by the energy level of said melody in the vicinity of the initial breakpoint.

19. A method according to claim 17, including eliminating from consideration initial breakpoints associated with confidence levels below a specified threshold, thereby identifying breakpoints in said melody.

20. A method according to claim 19, including estimating the pitch and beat duration of each said note between said breakpoints.

21. A method according to claim 1, wherein the melody is a voice-hummed melody composed of a series of uttered semi-vowels.

22. Apparatus for converting a digitized melody into a sequence of notes, comprising:

means for segmenting said melody into a series of frames;

means for computing a spectral energy distribution (SED) indicator for each frame; and

means for estimating initial breakpoints in said melody based on said SED, said notes being defined between adjacent initial breakpoints.

23. Apparatus according to claim 22, wherein the value of said SED indicator for a given frame is relatively large if an energy distribution associated with said frame is concentrated in one or more specified frequency bands.

24. Apparatus according to claim 23, including filtering said melody with a high pass filter prior to segmenting said melody into said frames.

25. Apparatus according to claim 24, wherein said energy distribution is determined from a normalized energy spectrum of said frame.

26. Apparatus according to claim 24, wherein said specified frequency band is the upper portion of a 0 to 4 kHz range.

27. A method for converting a digitized melody into a sequence of notes, comprising:

segmenting said melody into a series of frames;

computing the auto-correlation of each said frame;

estimating the pitch of each said frame based on (i) a pitch period corresponding to a shift where the auto-correlation coefficient associated with the frame is relatively large and (ii) the closeness of the pitch estimate to estimates in one or more adjacent frames; and

estimating breakpoints in said melody based on changes in said pitch estimates, said notes being defined between adjacent breakpoints.

28. A method according to claim 27, wherein said breakpoints are estimated based on a rate of change of said pitch estimates.

29. A method according to claim 27, including filtering said melody with a band pass filter prior to segmenting the melody into frames.

30. A method according to claim 27, including estimating the pitch of each note by selecting the average or median pitch of the frames falling within a pair of breakpoints.

31. A method according to claim 27, wherein the melody is a voice-hummed melody.

32. Another aspect of the invention provides a method for identifying breakpoints in a digitized melody, the method comprising:

segmenting the melody into a series of frames;

computing the auto-correlation of each frame;

estimating the pitch of each frame based on (i) a pitch period corresponding to a shift where the auto-correlation coefficient associated with the frame is relatively large and (ii) the closeness of the pitch estimate to estimates in one or more adjacent frames;

determining regions of said melody where pitch estimates are likely to be invalid; and

identifying said breakpoints in the melody based on transitions between frames having valid pitch estimates and transitions having invalid pitch estimates.

33. A method according to claim 32, wherein said breakpoints are estimated based on a rate of change of said pitch estimates.

34. A method according to claim 32, including filtering said melody with a band pass filter prior to segmenting the melody into frames.

35. A method according to claim 32, including estimating the pitch of each note by selecting the average or median pitch of the frames falling within a pair of breakpoints.

36. A method according to claim 32, wherein the melody is a voice-hummed melody.

37. Apparatus for converting a digitized melody into a sequence of notes, comprising:

means for segmenting said melody into a series of frames;

means for computing the auto-correlation of each said frame;

means for estimating the pitch of each said frame based on (i) a pitch period corresponding to a shift where the auto-correlation coefficient associated with the frame is relatively large and (ii) the closeness of the pitch estimate to estimates in one or more adjacent frames;

means for determining regions of said melody where pitch estimates are likely to be invalid; and

means for estimating breakpoints in said melody based on changes in said pitch estimates or transitions between frames having valid pitch estimates and frames having no pitch estimates, said notes being defined between adjacent breakpoints.

38. A method of retrieving at least one entry from a music database, wherein each said entry is associated with a sequence of pitches and beat durations, said method comprising:

receiving a digitized representation of an input melody;

identifying breakpoints in said melody in order to define notes therein, each said notes being delineated by adjacent breakpoints;

assigning a confidence level to each note or each breakpoint;

determining a pitch and beat duration for each note of said melody;

determining a score for each said entry based on a search which minimizes the cost of matching the pitches and beat durations of said melody and said entry, wherein said search considers at least one deletion or insertion error in a selected note of said melody and, in this event, penalizes the cost of matching based on the confidence level of the selected note or a breakpoint associated therewith; and

presenting said at least one entry to a user based on its score.

39. A method according to claim 38, wherein said pitches and beat durations are relative pitches and relative beat durations.

40. A method according to claim 38, wherein the cost of matching a given note X_i of said melody with a given note Y_j associated with said entry is:

$match\_\cos t(X_i, Y_j) = \alpha|YRF_j - XRF_j| + \beta|YRT_j - XRT_i|$, where YRF_j and YRT_j respectively represent the relative pitch and relative beat duration of the note associated with said entry; XRF_i and XRT_i respectively represent the relative pitch and relative beat duration of the note associated with said melody; and α and β are weights.

41. A method according to claim 38, wherein:

a confidence level is assigned to each note and each breakpoint; and

said search considers deletion and insertion errors for any given note of said melody and, in this event, penalizes the cost of matching based on the confidence level of the given note and the confidence level of a breakpoint associated with the given note.

42. A method according to claim 41, wherein:

X is a sequence of notes, X_i, of said melody, each X_i having components XRF_i, XRT_i $XICON_i$, and $XDCON_i$ which respectively represent the relative pitch, relative beat duration, confidence level of the breakpoint and confidence level of the note associated with said melody;

Y is a sequence of notes, Y_j, of said entry, each Y_j having components YRF_j and YRT_j which respectively represent the relative pitch and relative beat duration of the note associated with said entry;

X and **Y** form a matrix, and at a matching point (X_i, Y_j) said search seeks to identify a preceding set of notes $\{(X_{i-1-k}, Y_{j-1}), (X_{i-1}, Y_{j-1-k})\}$, $0 \le k \le \max_k$, which minimize a match cost defined as follows:

if k = 0, $\alpha|YRF_j - XRF_{u-1}| + \beta|YRT_{j-1} - XRT_{i-1}|$,

else if k > 0,

$$\alpha|YRF_{j-1} - XRF_{i-1-k}| + \beta|YRT_{j-1} - XRT_{i-1-k}| + \sum_{m=0}^{k-1}(\text{penalty for the } (m+1)^{th} \text{ insertion}) * XICON_{i-1-m} \text{ or}$$

$$\alpha|YRF_{j-1-k} - XRF_{i-1}| + \beta|YRT_{j-1-k} - XRT_{i-1}| + (\text{penalty for k deletions}) * XDCON_{i-1}$$

where α and β are weights.

43. Apparatus for retrieving at least one entry from a music database, wherein each said entry is associated with a sequence of pitches and beat durations, said apparatus comprising:

means for receiving a digitized representation of an input melody;

a melody-to-note conversion subsystem for identifying breakpoints in said melody in order to define notes therein, said subsystem determining a pitch and beat duration for each note of said melody and associating each note or each breakpoint with a confidence level;

a note-matching engine for determining a score for each said entry based on a search which minimizes the cost of matching the pitches and beat durations of said melody and said entry, wherein said search considers at least one deletion or insertion error in a selected note of said melody and, in this event, penalizes the cost of matching

based on the confidence level of the selected note or a breakpoint associated therewith; and

an output subsystem for presenting said at least one entry to a user based on its score.

44. A method of retrieving at least one entry from a music database, wherein each said entry is associated with a sequence of pitches and beat durations, said method comprising:

receiving a digitized representation of an input melody;

identifying breakpoints in said melody in order to define notes therein, each said notes being delineated by adjacent breakpoints;

associating a confidence level with each note pertaining to likelihood that said note contains a note insertion error;

determining a pitch and beat duration for each note of said melody;

determining a score for each said entry based on a search which minimizes the cost of matching the pitches and beat durations of said melody and said entry, wherein said search considers at least one insertion error in a selected note of said melody and, in this event, penalizes the cost of matching based on the confidence level associated with the selected note; and

presenting said at least one entry to a user based on its score.

45. A method of retrieving at least one entry from a music database, wherein each said entry is associated with a sequence of pitches and beat durations, said method comprising:

receiving a digitized representation of an input melody;

identifying breakpoints in said melody in order to define notes therein, each said notes being delineated by adjacent breakpoints;

associating a confidence level with each note pertaining to likelihood that said note contains a note deletion error;

determining a pitch and beat duration for each note of said melody;

determining a score for each said entry based on a search which minimizes the cost of matching the pitches and beat durations of said melody and said entry, wherein said search considers at least one deletion error in a selected note of said melody and, in this event, penalizes the cost of matching based on the confidence level associated with the selected note; and

presenting said at least one entry to a user based on its score.

46. A method for determining confidence levels for breakpoints or notes in a waveform representing a melody, the method comprising:

segmenting the waveform into a series of frames, wherein adjacent breakpoints encompass one or more sequential frames;

executing at least two of the following three steps,

(a) computing a spectral energy distribution (SED) indicator for each frame,

(b) estimating the pitch of each frame, and

(c) determining the energy level of each frame,

deriving the confidence levels based on at least two of the following three characteristics, (i) the SED indicator, (ii) changes in pitch, and (iii) the energy level.

47. A method according to claim 46, wherein the confidence level for a given breakpoint is computed as a weighted combination of at least two of three numbers, the first number based on the value of the SED indicator in the vicinity of the given breakpoint, the second number being based on a change in pitch in the frames before and after the given breakpoint, and the third number being based on the energy level of the frames in the immediate vicinity of the breakpoint.

48. A method according to claim 46, wherein the confidence level for a given note is computed as a weighted combination of at least two of three numbers, the first number based on the value of the SED indicator in the given note, the second number being based

on the variation in pitch in the given note, and the third number being based on the energy level of the frames in the given note.

49. A method for determining confidence levels for breakpoints or notes in a waveform representing a melody, the method comprising:

segmenting the waveform into a series of frames, wherein adjacent breakpoints encompass one or more sequential frames;

computing a spectral energy distribution (SED) indicator for each frame;

estimating the pitch of each frame; and

deriving the confidence levels based on the SED indicator and changes in pitch.

50. A method according to claim 49, wherein the confidence level for a given breakpoint is computed as a weighted combination of a first number based on the value of the SED indicator in the vicinity of the given breakpoint and a second number based on a change in pitch in the frames before and after the given breakpoint.

51. A method according to claim 49, wherein the confidence level for a given note is computed as a weighted combination of a first number based on the value of the SED indicator within the given note and a second number based on the variation in pitch within the given note.

52. A method according to claim 49, wherein the value of the SED indicator for a given frame is relatively large if an energy distribution associated with the frame is concentrated in one or more specified frequency bands.

53. A method according to claim 52, including filtering the melody with a high pass filter prior to segmenting the melody into frames.

54. A method according to claim 53, wherein the energy distribution is determined from a normalized energy spectrum of the frame.

55. A method according to claim 54, wherein the specified frequency band is in the upper portion of a 0-4kHz frequency range.

56. A method for determining confidence levels for breakpoints or notes in a waveform representing a melody, the method comprising:

segmenting the waveform into a series of frames, wherein adjacent breakpoints encompass one or more sequential frames;

computing a spectral energy distribution (SED) indicator for each frame;

determining the energy level of each frame; and

deriving the confidence levels based on the SED indicator and the energy level.

57. A method according to claim 56, wherein the confidence level for a given break point is computed as a weighted combination of a first number based on the value of the SED indicator in the vicinity of the given breakpoint and a second number based on the energy level of the frame in the immediate vicinity of the breakpoint.

58. A method according to claim 56, wherein the confidence level for a given note is computed as a weighted combination of a first number based on the value of the SED indicator in given note and a second number based on the energy level of the frames in the given note.

59. A method according to claim 56, wherein the value of the SED indicator for a given frame is relatively large if an energy distribution associated with the frame is concentrated in one or more specified frequency bands.

60. A method according to claim 59, including filtering the melody with a high pass filter prior to segmenting the melody into frames.

61. A method according to claim 60, wherein the energy distribution is determined from a normalized energy spectrum of the frame.

62. A method according to claim 61, wherein the specified frequency band is the upper portion of a 0-4kHz frequency range.